Junseok Park

junseokpark.kr@gmail.com • junseokpark.github.io • github.com/junseokpark

A computational scientist, data scientist, or bioinformatician with expertise in computer science, skilled in developing machine learning models, designing large-scale data processing pipelines, and performing statistical analyses of next-generation sequencing and multi-omic experiments from patient cohorts—ranging from single-cell to DNA and RNA sequence samples. Leveraging strong analytic thinking, I have contributed to projects ranging from cancer immunology and amyotrophic lateral sclerosis to dementia and neurodegeneration in ataxia-telangiectasia.

Education

2016–2020 **PhD** program in **Bioinformatics** – KAIST, Daejeon, South Korea
2014–2016 **MSc** in **Bioinformatics** – KAIST, Daejeon, South Korea
2000–2007 **BSc** in Computer Engineering – Chonnam National University, Gwangju, South Korea

Skills

Tool developments: rTea, GenomeFlow, CORUS, CLIPS
Computational skills: Deep learning, Machine learning, Statistical and Network Analysis
Deeplearning frameworks: TensorFlow, Pytorch and Keras
Programming languages: Python, R, Scala, C++, Java, JavaScript, Shell Scripts, SQL and noSQL
Computing platforms: On-premises HPC (Slurm/Qsub), AWS, GCP and Terra.bio
Workflow development: WDL, Snakemake, Airflow and Kubernetes
TA Courses: BiS301 Bioengineering, BiS732 Bio-Network, BiS332 Bio Data Mining, BiS232 Bio-Data Structure
Languages: Korean (fluent), English (fluent), Japanese (intermediate)

Work Experience

Aug. 2020 – present **Lee Lab** – Boston Children's Hospital and Harvard Medical School, Boston, MA, USA
Research Fellow (PI: Dr. Alice Eunjung Lee)

- Conducting statistical genetics research or biostatistics techniques to support research hypotheses derived from high-throughput DNA, RNA, and single-cell sequencing data.
- Developing computational pipelines to identify disease-causing genetic variants.
- Coordinating an interdisciplinary team of researchers to develop and apply deep learning and machine learning methods in computational biology projects.
- Performing comprehensive data analysis to interpret complex biological datasets and derive meaningful insights for disease-related research.
- Ensuring quality control of the sample to maintain the reliability of genomic analyses.
- Leveraging cloud computing resources for processing large-scale datasets efficiently.

Jan. 2012 – Feb 2014 **Korea Research Institute of Chemical Technology (KRICT)** – Daejeon, Republic of Korea
Technical Manager

- Collaborated in a team-oriented environment to develop a chemical reagent management system, enhancing laboratory efficiency by optimizing reagent usage workflows.
- Leveraged Oracle DB, HTML5, and JavaScript engines to develop the system.

Nov 2011 – Dec. 2011 **Seoul National University Hospital** – Seoul, Republic of Korea
Planning Manager

- Integrated medical records (EMR/EHR/PHR) to streamline medical record management and develop a hospital management system, using interdisciplinary communication skills

Apr. 2011 – Nov. 2011 **Korea BIO-IT Foundry Gwang-ju Centre** – Gwangju, Republic of Korea
Researcher (PI: Dr. Yonggwan Won)

- Designed and developed a high-speed additive reagent injection and inspection system, demonstrating effective time management skills.
- Utilized a Support Vector Machine algorithm to enhance precision in the biotech system.

Jan. 2007 – Mar. 2010 **SKTelecom** – Gwangju/Seoul, Republic of Korea
Manager

- Worked independently with sub-contractors to optimize and enhance the performance of WCDMA networks, improving reliability and benefiting approximately 12 million customers.
- Applied management skills and a strong technical background to develop customized C++ program blocks for the Mobile Switching Center and Software Data Network.

Mar. 2001 – Apr. 2003 **Republic of Korea Army** – 31 Division, Republic of Korea
Sergeant, Honorable discharge

- Developed ocean protection radar systems and networks along the Korean coast, motivated to enhance coastal surveillance and maritime safety.
- Utilized Visual C++ 6.0 to design systems capable of recording and analyzing critical data.

Select Publications

- Boram Lee*, **Junseok Park***, Adam Voshall, Yangmin Gan, Eduardo Maury, Yeeok Kang et al. *Pan-cancer analysis reveals multifaceted roles of retrotransposon-fusion RNAs.* Nature Communication (2024): In revision.
  10.1101/2023.10.16.562422, github.com/junseokpark/rtea
- **Junseok Park**, Eduardo Maury, Changhoon Oh, Donghoon Shin, Danielle Denisko, Eunjung Alice Lee. *Genomic data processing with GenomeFlow.* Accepted to BMC Bioinformatics (2024).
  github.com/junseokpark/genomeflow
- Miaomiao Tan*, Zhinan Lin*, Zhuofu Chen, **Junseok Park**, Ziting He, Haonan Zhou et al. *Image-based DNA Sequencing Encoding for Detecting Low-Mosaicism Somatic Mobile Element Insertions.* Nature Method (2024): In revision. 10.1101/2024.11.07.619809
- Jeffrey J. Widrick, Matthias R. Lambert, Felipe de Souza Leite, Youngsook Lucy Jung, **Junseok Park**, James R. Conner et al. *Kinematic phenotyping of dystrophic zebrafish larvae*, Science advances (2024): In revision. 10.1101/2024.12.05.627004
- Lai, Jenny, Didem Demirbas, Junho Kim, Ailsa M. Jeffries, Allie Tolles, **Junseok Park** et al. *ATM-deficiency-induced microglial activation promotes neurodegeneration in ataxia-telangiectasia.* Cell reports (2024): 43, no. 1. 10.1016/j.celrep.2023.113622
- Zinan Zhou, Junho Kim, August Yue Huang, Matthew Nolan, **Junseok Park**, Ryan Doan et al. *Somatic Mosaicism in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Reveals Widespread Degeneration from Focal Mutations.* [Preprint] bioRxiv (2023). 10.1101/2023.11.30.569436
- **Junseok Park**, Kwangmin Kim, Seongkuk Park, Woochang Hwang, Sunyong Yoo, Gwansu Yi, Doheon Lee. *An interactive retrieval system for clinical trial studies with context-dependent protocol elements.* PloS one (2020): 15.9:e0238290. 10.1371/journal.pone.0238290, github.com/junseokpark/clips
- **Junseok Park**, Seongkuk Park, Kwangmin Kim, Gwangmin Kim, Jaegyun Jung, Sunyong Yoo et al. *Reliable Data Collection in Participatory Trials to Assess Digital Healthcare Applications.* IEEE Access (2020): 79472-79490. 10.1109/ACCESS.2020.2985122, github.com/junseokpark/corus
- **Junseok Park**, Kwangmin Kim, Woochang Hwang, and Doheon Lee. *Concept embedding to measure semantic relatedness for biomedical information ontologies.* Journal of biomedical informatics (2019): 94:103182. 10.1016/j.jbi.2019.103182

Patents

Jan 2022 Method and Apparatus for Performance Evaluating of Healthcare Applications
KR20210081545A, https://patents.google.com/patent/KR20210081545A

Jan 2020 Method and Apparatus for Data Managing for Clinical Trial
KR20190094729A, https://patents.google.com/patent/KR20190094729A

Jul 2017 A Method for Searching Co-Occurrence Based on Co-Operational Formation
KR20160149619A, https://patents.google.com/patent/KR20160149619A

Certificates

2011–Present **Network Management** – Information & Communication Qualification Association of Korea
2011–Present **Engineer Information Processing** – Human Resources Development Service of Korea
2009–2010 **Six Sigma Green Belt** – SKTelecom, Seoul, Korea
2007–2009 **Cisco Certified Network Associate** – CISCO, Seoul, Korea

*Co-first authors